

Improvement of mitochondria extract from *Saccharomyces cerevisiae*
characterization in shotgun proteomics using sheathless capillary
electrophoresis coupled to tandem mass spectrometry

Marianne Ibrahim¹, Rabah Gahoual¹, Ludovic Enkler², Hubert Becker², Johana Chicher³,
Philippe Hammann³, Yannis-Nicolas François¹, Lauriane Kuhn³, Emmanuelle Leize-Wagner¹

Author information:

¹ Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UDS-
CNRS UMR 7140, Université de Strasbourg, 67008 Strasbourg, France.

² Unité Mixte de Recherche 7156 Génétique Moléculaire Génomique Microbiologie, Centre
National de la Recherche Scientifique, Université de Strasbourg, 67084 Strasbourg, France.

³ Plateforme Protéomique Strasbourg-Esplanade, Institut de Biologie Moléculaire et
Cellulaire, FRC 1589, Centre National de la Recherche Scientifique, Université de
Strasbourg, 67084 Strasbourg, France.

Corresponding author: Yannis-N. François, yfrancois@unistra.fr

ABSTRACT

In this work, we describe the characterization of a quantity-limited sample (100ng) of yeast mitochondria by shotgun bottom-up proteomics. Sample characterization was carried out by sheathless capillary electrophoresis, equipped with a high sensitivity porous tip and coupled to tandem mass spectrometry (CESI-MS/MS) and concomitantly with a state-of-art nano flow liquid chromatography coupled to a similar mass spectrometry (MS) system (nanoLC-MS/MS). With single injections, both nanoLC-MS/MS and CESI-MS/MS 60 min-long separation experiments allowed to identify 271 proteins (976 unique peptides) and 300 proteins (1765 unique peptides) respectively, demonstrating a significant specificity and complementarity in identification depending on the physicochemical separation employed. Such complementary, maximizing the number of analytes detected, presents a powerful tool to deepen a biological sample's proteomic characterization. A comprehensive study of the specificity provided by each separating technique was also performed using the different properties of the identified peptides: molecular weight (MW), mass-to-charge ratio (m/z), isoelectric point (pI), sequence coverage or MS/MS spectral quality enabled to determine the contribution of each separation. For example, CESI-MS/MS enables to identify larger peptides and eases the detection of those having extreme pI without impairing spectral quality. The addition of peptides, and therefore proteins identified by both techniques allowed to increase significantly the sequence coverages and then the confidence of characterization. In this study, we also demonstrated that the 2 yeast enolase isoenzymes were both characterized in the CESI-MS/MS dataset. The observation of discriminant proteotypic peptides is facilitated when a high number of precursors with high-quality MS/MS spectra are generated.

Introduction

In recent years, mass spectrometry-based proteomics was recognized as one of the best tools for identifying proteins with new functions, mapping their interactions in a cellular context and discovering new biomarkers for medical research ¹. In order to maximize the number of analytes successfully ionized, detected and identified in mass spectrometry (MS), a physicochemical separation prior to MS analysis is often applied. Currently, liquid chromatography (LC) is the most widely used separation techniques that can be coupled on-line with a mass spectrometer ². In a classical nanoLC coupled to tandem mass spectrometry (nanoLC-MS/MS) experiment, which is the most usual method for proteomic application, analytes are eluted from a reverse-phase column by increasing the organic content of the mobile phase to separate peptides by hydrophobicity. Capillary Zone Electrophoresis (CZE) is another separating technique that has been successfully coupled to electrospray ionization ³: with the use of an electrical field, ions are separated based on their electrophoretic mobility which is dependent upon the charge of the molecule and the analyte's hydrodynamic radius ⁴. Capillary electrophoresis (CE) is gaining more and more interest mostly because of recent technical improvements regarding capillary zone electrophoresis–mass spectrometry (CZE-MS) coupling that allow the use of lower flow rates resulting in improved sensitivity ^{5,6} and also because ion suppression phenomenon is largely reduced compared to a nanoLC-MS analysis ⁷⁻⁹. Recently, CE–MS has emerged as a highly efficient separation technique ¹⁰ and is considered as a powerful technique for the analysis of peptides and proteins ^{4,11-22}.

Taking into considerations these observations, an ideal workflow would be the combination of orthogonal separative techniques based on different physiochemical principles. Indeed, as a significant fraction of proteins can escape detection in individual separation approaches²³, crossing over different techniques and their respective benefits would minimize that problem. Consecutive CZE-MS/MS and nanoLC-MS/MS analysis campaigns can be easily imagined on the same samples as further discussed in this study with the example of a quantity-limited sample from a yeast mitochondrial extract.

In this study, we used CZE-MS/MS as well as a classical nanoLC configuration to explore the *S. cerevisiae* mitochondrial proteome. CZE-MS/MS experiments were performed using a recently introduced CE-MS interface referred as CESI-MS. New instrumental approaches are often developed and validated on model systems such as the baker's yeast *Saccharomyces cerevisiae*, mainly because many essential processes are conserved between yeast and other organisms of interest like humans²⁴. Numerous benefits derive from the fact that yeast is one of the simplest eukaryotes²⁵. Because of its importance in molecular biology, it was the first eukaryotic organism for which the genome sequence was completed²⁶. Mitochondrion is a complex intracellular organelle, crucial for numerous cellular functions like normal cell metabolism²⁷, cellular energetic²⁸, maintenance of ion homeostasis²⁹ and programmed cell death^{30,31}. To achieve the best results, a highly purified mitochondrial preparation is mandatory^{32,33}. To understand the role of this complex organelle especially in disease, it is important to extensively characterize its protein composition: identifying the whole set of resident proteins within a complex organelle remains a major challenge in cell biology even if innovative technologies are emerging year after year. Previous proteomic analyses on purified mitochondria resulted in a repertoire of 1000 to 1500 different proteins for that organelle, in either yeast or human samples^{34,35}.

In the current work, the aim is to demonstrate the improvement in protein identification by shotgun bottom-up proteomics of mitochondria extract from *Saccharomyces cerevisiae* by using two different physicochemical separations prior to the MS analysis. The study is carried out on a limited quantity (100ng) of sample, roughly representing 1000 proteins. Identifications obtained by CESI-MS/MS were compared to those achieved in nanoLC-MS/MS. Specificity of each separating technique was determined observing different physicochemical properties of the identified peptides: molecular weight (MW), mass-to-charge ratio (m/z), isoelectric point (pI), sequence coverage as well as MS/MS spectral quality. Moreover, a cellular localization analysis of all identified proteins was also performed. Biological significance was finally investigated by comparing the protein identifications from this study to previously published studies. In addition, emphasis was placed onto a challenging area of proteomics research which is isoforms characterization. Taking together, the results presented in this study allowed us to assess the gain of information which can be achieved by two complementary and orthogonal techniques, in terms of peptide metrics, sequence coverage, spectrum quality and isoforms characterization.

Experimental

Materials and reagents.

Chemicals used were of analytical grade or high purity grade and purchased from Sigma-Aldrich (Saint Louis, MO, USA). Water used to prepare buffers, mobile phases and sample solutions was obtained using an ELGA purelab UHQ PS water purification system (Bucks, UK). Dithiothreitol (DTT) and iodoacetamide (IAA) were purchased from Sigma (St. Louis, MO, USA). Trypsin sequencing grade was obtained from Promega (Madison, WI USA). All other reagents and plastic ware were obtained from commercial sources.

Isolation of Mitochondria from *S. cerevisiae*

Yeast cells were grown in Yeast Peptone Dextrose (YPD) at 30°C up to OD_{600nm}=1.2. Mitochondria were isolated as detailed in the Supplemental Information section. Briefly, spheroplasts have been first generated by the enzymatic digestion of the cells wall by Zymolase 20T (Euromedex) and further broken by homogenization in a glass-Teflon potter. After the removal of cell debris and nuclei by centrifugation (1,500×g, 5 min at 4°C), mitochondria were recovered from the supernatant by additional centrifugations at 12,000×g for 15 min at 4°C, to give a final concentration of 5 mg/ ml. Mitochondrial fraction was treated by 10 strokes in a glass-Teflon potter and loaded onto a four-step sucrose gradient, from which purified mitochondria were recovered from the 60 %/ 32 % interface.

In solution protein digestion

10µg of protein were diluted in 75µL ammonium bicarbonate buffer (50mM) and incubated with 5µL DTT 0.1M for protein denaturation (10min, 95°C). After cooling down, 10µL IAA 0.1M were added and the samples were incubated for 20min at room temperature in the dark to allow alkylation of reduced cysteine residues. For protein digestion, 500ng (1:20) trypsin (Promega, V5111) was added and sample was incubated overnight at room temperature. For nano-LC/MSMS analysis, 3µg of the clear supernatant were transferred into glass inserts and evaporated to dryness using a miVac DNA concentrator (Genevac, NY, USA). The sample was either reconstituted in 15µL formic acid (0.1% v/v) for nanoLC-MS/MS analyses or in ammonium acetate 50 mM (pH 4.0) for CESI-MS/MS analyses.

NanoLiquid Chromatography

Samples were transferred into glass inserts, compatible with the LC autosampler system (ultra nanoLC-2D plus, Eksigent, UK). Five microliters of each sample were loaded on a

nanoFlexcHiP module consisting of a Trap-and-elute jumper (fluidics configuration design), a ChromXP-C18 trap cHiP column (0.5cm x 200µm i.d., 3µm, 120Å, Eksigent) and a ChromXP-C18 analytical cHiP column (15cm x 75µm i.d., 3µm, 120Å, Eksigent). The separation method consisted in a 60 min run at a flow rate of 300 nL/min using a gradient of two solvents: A (99.9% water: 0.1% formic acid) and B (99.9% acetonitrile: 0.1% formic acid). After a 10 min step to trap the peptides on the pre-concentration column at 20µL/min, they were eluted from the analytical column as follows: 0-35 min, 5-40% B; 35-36 min, 40-80% B; 36-40 min, 80% B; 40-41 min, 80-5% B; 41-60 min, 5% B.

Capillary Electrophoresis

The CE experiments were performed with a CESI8000 capillary electrophoresis system from Sciex separations (Brea, CA) equipped with a temperature controlled autosampler and a power supply able to deliver up to 30 kV. Bare fused-silica capillaries (total length 91 cm; 30µm i.d.) with characteristic porous tip on its final 3 cm supplied by Sciex separation allowed electric contact between both electrodes of the CE system. New capillaries were flushed at 75 psi (5.17 bar) for 10min with methanol, then 10min with 0.1M sodium hydroxide, followed 10min with 0.1M hydrochloric acid and water for 20min. Finally, the capillary was flushed 10min at 75 psi with BGE which was acetic acid 10%. Hydrodynamic injection (410 mbar for 1min) corresponding to a total volume of 55 nL of sample injected was used. Separations were performed using an electric field of +219.8 V/cm⁻¹.

Mass spectrometry

The LC and CESI systems were coupled to a 5600Triple TOFTM system (AB Sciex, CA, USA) with nanospray source operating in positive ESI mode. Interchangeability of the LC and CESI systems is easy and takes only 20 min. The mass spectrometer was operated in the

IDA mode (Information-Dependent Acquisition) and was externally calibrated each 6 hours with 25fmol of an *E.coli* Beta-galactosidase digest. This standard also allowed to regularly check the performances and reproducibility of the mass spectrometer. For the nanoLC-MS/MS, the following ESI source parameters were used: ion spray voltage, 2.3 kV; ion source heater, 150 °C; curtain gas, 22 psi and ion source gas 1,5 psi. While for the CESI-MS/MS and due to the nanospray properties, the ESI source parameters were: ion spray voltage, 1.75 kV (1.0 to 4.0 mm between the MS inlet and capillary tip); ion source heater, 75 °C; curtain gas, 5 psi (flow rate < 100 nL/min) and ion source gas 1, 0 psi. The MS survey scan was acquired over a 400-1250 *m/z* mass range with a 250ms accumulation time; whereas the product ion experiments mass range was extended to 100-2000 *m/z* with a 60ms accumulation time. The duty cycle time was therefore maintained below 2s with up to 30 precursor ions selected on the MS survey scan to undergo CID fragmentation (Collision Induced Dissociation), while respecting the following criteria: intensity greater than 150 counts per second, charge state between 2+ and 5+, not present in the dynamic exclusion list, exclusion after a period of 15s and 2 consecutive MS/MS spectra, and use of a rolling collision energy. Data were acquired with Analyst 1.6 software (AB Sciex) and the resulting .wiff files were primarily evaluated with PeakView 1.2 software (AB Sciex).

Data Analysis

For protein identification, the MS/MS raw data files were first converted to mascot generic format (mgf) using the executable MS-DataConverter.exe (AB Sciex). MS data processing was then performed with Proteinscape software (version 3.1, Bruker Daltonics). We used the Mascot algorithm (v2.2, Matrix Science, UK) for database searching, configured with the following parameters: (a) database: Swiss-Prot with *Baker's Yeast* taxonomy (version 2013-01-09, 18860 sequences), (b) enzyme: Trypsin/P, (c) maximum missed cleavages: 3, (d)

variable modifications: acetylation (Protein N-term), oxidation (Met), carbamidomethylation (Cys), (e) peptide mass tolerance: 30 ppm, (f) MS/MS tolerance: 0.5 Da, (g) Instrument: ESI-Quad-TOF. To evaluate the false-positive rate of each analysis, we generated a reversed yeast database using a Pearl script (makeDecoyDB.pl, Bruker): the fusion of this decoy database with the original forward yeast database allowed Proteinscape to validate the data at FDR<1%. Moreover, redundancy was taken into account by Proteinscape (grouping of proteins sharing exactly the same set of peptides) and a manual inspection of the MS/MS fragmentation spectra was done for proteins identified with 1 peptide and a Mascot Score below 50. The same MS/MS raw data files were then submitted to a second database search algorithm, (Paragon, AB Sciex) with the following parameters: (a) database: the same Baker's Yeast taxonomy as described previously for Mascot, (b) enzyme: Trypsin, (c) Cys alkylation: Iodoacetamide, (d) ID focus: biological modifications, (e) search effort: thorough ID, (f) confidence p-value: 0.05, (g) Instrument: Triple-TOF 5600.

To facilitate the biological interpretation of the protein identification lists, the PANTHER classification system (<http://pantherdb.org/>) has been applied in a first instance on both nanoLC-MS/MS and CESI-MS/MS datasets. A second classification system named iLoc-Euk (<http://www.jci-bioinfo.cn/iLoc-Euk>) was then used to assess the subcellular localization of each identified protein. Combining both prediction softwares during the final data mining step allowed to strengthen the confidence related to each protein biological significance. Parameters used for both systems are detailed in the Supplemental Information section.

Results

Peptide metrics to highlight CESI-MS/MS and nanoLC-MS/MS approaches.

Consecutively to the development of the ESI ionization source and its application to study biological analytes in MS, technical developments have been achieved in order to

perform a separation prior to the MS analysis. Performing a separation before the ionization process allows to analyze samples with an increased complexity and tends to enhance the sensitivity of the signal by limiting the competition effect during ESI ionization³⁶. Interfaces allowing CE-MS hyphenation were therefore developed as CE is well suited for the separation of biological analytes such as peptides or proteins while providing high efficiency separation due to the absence of stationary phase. In this work, a recently introduced sheathless CE-MS interface (CESI) has been used. A detailed description of the system has been provided in the literature by Haselberg et al.¹³. Recently, following the improvement of shotgun proteomics with recent mass spectrometer, different research groups have described shotgun proteomic experiments performed by sheathless CE-MS using commercially available or custom-made interfaces^{37,38}. One of the major goals of this study was in a first instance to identify a great number of proteins from a relatively low amount of yeast mitochondrial extract. For that purpose, we used either the state-of-art shotgun proteomic configuration with a RP-nanoLC-MS/MS interface, or the promising CESI-MS/MS interface³⁹⁻⁴¹. ESI parameters for NanoLC and CESI systems are defined in order to be compatible with nanoESI flowrate (300 nL/min for nanoLC and <100 nL/min for CESI) and totally comparable in terms of performance (See Experimental section). 100ng of material, a quantity which can be typically obtained after a sub-cellular fractionation process on a wide variety of organisms, were injected on both couplings. When considering studies published during the last 5 years, we can notice that starting sample amounts ranging from 1ng to 400ng were often considered, with an average at 100ng (Supplemental Table S1). Moreover, as stated by Sun et al in their review published in 2014 (Sun et al., 2014), the vast majority of proteomic studies using a nanoLC-MS/MS approach are typically starting with biological samples quantities in the milligram or microgram range. That's why CESI-MS/MS approach appears as a good alternative when the starting material is limited and falling into the midnanogram range.

To obtain a simple picture from the two complex MS datasets generated, heat maps were reconstructed using PeakView software. The mass/charge ratio of the peptides (Da) is plotted against the LC elution time or the CZE migration time (Figures 1-A and 1-B). A nanoLC-MS/MS heat map, like the one we obtained for the injection of 100ng of a yeast mitochondrial extract, is generally characterized by a wide cloud of dots, each of them representing a unique MS/MS spectrum related to the fragmentation of a given precursor ion (Figure 1-A). This cloud of dots means that LC conditions are regularly eluting peptides from the reverse-phase column, without introducing any additional phenomenon except the fact that hydrophobic peptides will elute latter than hydrophilic peptides. In contrast, the heat map obtained with the CESI-MS/MS interface displays a very different picture: distinct and successive lines or “strikes” of analyzed peptides can be drawn on the graph (Figure 1-B). This pattern can be explained by the CZE migration process itself as it depends not only on the charge in solution of the analytes but also on their mass-to-charge ratios.

A yeast UniProtKB database was used to identify peptides and proteins and a maximum false-positive rate of 1% was applied to validate both MS datasets. As shown in the Venn diagram (Figure 1-C), 222 proteins were identified with either CESI-MS/MS or nanoLC-MS/MS analysis. In addition, 78 proteins were only identified by CESI-MS/MS while 49 other proteins were only identified by nanoLC-MS/MS. Proteins identified by both approaches represent the majority of the proteins identified in the two nanoLC-MSMS and CESI-MSMS analyses, 82% and 74% respectively. This complementarity observed between both couplings also agrees with previous reports ^{37,42}. Interestingly, CESI-MS/MS coupling allowed to identify 3 times more additional peptides compared to the classical nanoLC-MS/MS configuration (412 versus 1201 peptides), providing already a positive insight into the power of CESI-MS/MS for obtaining good sequence coverages. It must be emphasized that the number of identified proteins for each approach was obtained with a single injection

and is not the sum of multiple technical repeats. Among the 222 proteins identified commonly by both CESI-MS/MS and nanoLC-MS/MS, half of them are characterized by shared peptides, as well as additional peptides identified specifically by CE or LC (scenario #2 = EQUIV in Supplemental Figure S1): this scenario leads all the same to an equivalent result which is the identification of a given protein in the biological sample. Interestingly, this observation means that the two orthogonal methods of separation used in this study were identifying peptides bearing the same biophysical properties (564 common peptides, representing 26% of the whole dataset) but that there are more additional peptides only identified by the CE approach (1201) compared to the LC approach (412). This observation is also concordant with the fact that 41 proteins identified by LC are also identified by CE but with a higher number of peptides (scenario #3 = CE+ in Supplemental Figure S1). To note, 25% of the 222 common proteins are characterized by a completely different set of peptides (scenario #5 = DIFF in Supplemental Figure S1). This latest observation is nicely showing the usefulness of combining orthogonal separative methods before the mass spectrometry analysis.

To evaluate each type of coupling for identification of peptides bearing specific biophysical properties, we next represented the percentage of observed peptides obtained by each separation against their molecular weights (MW, Figure 2-A), mass-to-charge ratios (m/z , Figure 2-B), isoelectric points (pI, Figure 2-C) and Mascot scores (Figure 2-D). The most evident trend arises from the molecular weight distribution pattern: very large peptides were mostly detected with CESI-MS/MS rather than nanoLC-MS/MS. Indeed, only 10 peptides observed by nanoLC-MS/MS have a molecular weight above 2800 Da, whereas this number is reaching 466 peptides in the CESI-MS/MS injection: this might be explained by their irreversible adsorption to either the preconcentration or the chromatographic column, or the difficulty to correctly elute these large peptides. The detection of large peptides by CESI-

MS/MS is also a phenomenon visible in the medium mass range: the apexes of both MW distributions are separated by 400 Da, with the CESI-MS/MS data displaying the highest apex and the wider distribution. For example, in the medium mass range from 751 to 1250 Da, 52% of peptides were detected with CESI-MS/MS, *versus* 10% with nanoLC-MS/MS. On the contrary, in the low mass range from 400 to 750 Da, 90% of peptides were detected with nanoLC-MS/MS, *versus* 48% with CESI-MS/MS (Figure 2-A). The same results are obtained when we look at mass-to-charge ratios of the peptides, which are more widely used instead of the global molecular weight (Figure 2-B). In this study, no enrichment towards low MW peptides was observed in the CESI-MS/MS dataset. Interestingly, some other groups also highlighted, on another biological system, the fact that CESI-MS/MS allows to elute peptides of low MW¹⁸. These peptides are often composed of only a few amino acids, thereby giving them a hydrophilic property. This type of peptides is also more difficult to observe with nanoLC-MS/MS, as they are weakly bound or even unretained on reverse-phase stationary phases.

Another important feature to monitor is the pI value from the identified peptides. Here, we used the « Compute MW/pI » tool from ExPASy to obtain the isoelectric point values for all the observed peptides, resulting in a bi-modal distribution with unevenly distributed peptidic pI values across the pH scale, with a gap at pH 7-8, and the majority of peptides clustering at pH 4-6 (Figure 2-C). It can be noted that the pI distribution obtained on this *S. cerevisiae* mitochondrial extract is in agreement with previously published data like theoretical 2D-gels reconstructed by Knight et al.⁴³ on a wide variety of organisms including the total yeast proteome as well as sub-cellular yeast fractions. Results on pI distribution obtained in this study on a mitochondrial yeast extract are relatively similar. Slight differences can be observed all the same for the extreme pI values, especially extremely basic peptides (pI>10) or peptides with more acidic property (pI<4), as already shown by other research groups³⁷.

This phenomenon is potentially a result of the decreased ion suppression at very low flow rates and could be of particular interest for phosphopeptides detection without any enrichment techniques⁴².

Discussions

Increasing sequence coverage versus spectrum quality.

Maximal sequence coverage is recommended to increase protein identification confidence but coverage should not be won to the detriment of the spectral quality. As this parameter can be evaluated with the Mascot Score obtained for each spectrum, we next represented the Mascot Score distributions obtained on the CESI-MS/MS and nanoLC-MS/MS injections of 100ng mitochondrial extract (Figure 2-D). Interestingly, there are more peptides with a score lower than 61 identified by the nanoLC-MS/MS approach. The CESI-MS/MS coupling provides more peptides with a score upper than 61: these high quality MS/MS spectra are representing 39.8% of the whole CESI dataset and 27.7% of the whole LC dataset. As for the median Mascot Score, it shows a slight increase from 48.3 to 52.8, for the LC or CE separations respectively (Supplemental Figure S4-A). A wider distribution is also observed for the CESI-MS/MS dataset, and this observation is more noticeable when considering the sub-set of peptides composed of the LC- and CE-specific features (Supplemental Figure S4-A). This Mascot Scores comparison is thus reflecting the overall good spectral quality obtained for both types of coupling.

As Wang et al³⁹ did previously while using the Sequest XCorr, we decided to use the Mascot score, which is probably the most widely-used database search algorithm by the proteomic community. Moreover, to strengthen these data, we also used a second dataset search algorithm, namely Paragon (ProteinPilot, AB Sciex): relevant observations can thus be conducted by combining 2 different algorithms, as recommended by proteomics guidelines.

Paragon algorithm returns the percentage of MS/MS spectra above a given confidence threshold: we can observe that there are always more fragmentation spectra above this threshold in the CESI-MS/MS dataset compared to the nanoLC-MS/MS dataset, whether the confidence threshold is set (up to 99%, Supplemental Figure S4-C).

There is a particular interest in the development of alternative technologies that could improve sequence coverages, to secure more confident identifications, especially for low-abundant proteins often characterized by a single or a few peptides. To study the usefulness of CESI-MS/MS to increase sequence coverage, we first ranked the identified proteins according to the total number of MS/MS fragmentation spectra matching to each of them. While plotting the resulting Spectral Count value as a function of the protein rank, we were able to select a panel of 5 different proteins which were homogeneously distributed all along the dynamic range from both datasets (Figure 3-A). When specifically inspecting the peptides distribution from this sub-set of proteins, we can observe that sequence coverage is indeed increased in the CESI-MS/MS analysis compared to the nanoLC-MS/MS analysis (Figure 3-B). While the effect on the sequence coverage increase is drastically observable on high-abundant proteins like PMA1, ODP4 and ADH1, a more reasonable impact is observed for mid- and low-abundant proteins like COX4 and ISD11. We further considered the whole set of proteins and thus represented the distribution of the global protein sequence coverage depending on the type of coupling (Figure 3-C) and observed the same trend. The example of the yeast plasma membrane ATPase PMA1 can illustrate the sequence coverage increase that was observed in this yeast mitochondrial dataset (Supplemental Figure S3-A). 24.8% sequence coverage was reported for PMA1 by nanoLC-MS/MS with 21 identified peptides, while sequence coverage from CESI-MS/MS was 43.5% with 45 identified peptides (Supplemental Figure S3-A). Interestingly, the 21 peptides identified by nanoLC-MS/MS were all identified by CESI-MS/MS (annotated in red on Supplemental Figure S3-A). Moreover, among the additional 24

peptides identified only by CESI-MS/MS, it can be noted that 7 of them have a large mass-to-charge ratio ($m/z > 1000$, highlighted in yellow on Supplemental Figure S3-A): these 5 unique peptides at [175-215], [216-252], [386-414], [482-508] and [584-615] are already representing 17.7% of the whole PMA1 sequence (17678 Da over 99619 Da for the full length protein). For the 21 peptides identified by both separative techniques, 17 of them have a CE Mascot Score higher than the corresponding LC Mascot Score (highlighted in grey in Supplemental Figure S3-A). To note, the MS/MS fragmentation spectra of the 5 largest peptides are of good quality as suggested by their average Mascot Score (84.6, Supplemental Figure S4-B) and have acidic pI values (average pI = 4.25). Such very large peptides are heavily contributing to increase the total sequence coverage. Interestingly, as suggested by other works, CESI-MS/MS is also able to identify small peptides that aren't by nanoLC-MS/MS. For example, the C-terminal part from PMA1 was covered in the CESI-MS/MS analysis with the doubly-charged peptide [910-918] being sequenced at $m/z=529.7611$, as well as 4 other small peptides (highlighted in grey on Supplemental Figure S3-A, peptides [272-278], [380-385], [429-435] and [436-442]). Similarly, when investigating the 4 other proteins distributed all along the dynamic range from the yeast mitochondrial sample (Figure 3-A), the same sequence coverage increase is observed (Supplemental Figures S3-B to S3-E).

***S.cerevisiae* mitochondrial proteins identification**

Biological interpretation of the data is of paramount importance for assessing the strength of mass spectrometric and proteomic results, and validation using orthogonal techniques is often requested to publish MS data. In this study, a focus was established on the yeast mitochondrial sub-proteome. Indeed, the determination of a protein's localization is useful for biologists because it is linked to its cellular function⁴⁴. It can also pinpoint some molecular functions to specific organelles⁴⁵. Thus, in order to perform a high-throughput cellular localization analysis of all identified proteins, two protein classification systems were

used: PANTHER⁴⁶ and iLoc-Euk⁴⁷. With the nanoLC-MS/MS approach (Figure 4-A), 55.3% of proteins have been identified and clustered as mitochondrial proteins while 53% of them were classified as mitochondrial proteins in the CESI-MS/MS dataset (Figure 4-B). In Figure 4-B, we can see that 127 mitochondrial proteins overlap between the two methods. Among the 349 proteins identified with both methods, 52.1% are known mitochondrial proteins, while the rest of them have been reported to be located into other subcellular compartments.

Identifying 100% of the proteins of an entire organelle from only one proteomic experiment remains a big challenge. Some proteins may escape from detection because of the technology used to separate the analytes upstream from the mass spectrometer itself: this effect is often emphasized when analytes are of extreme biophysical properties (like hydrophilic or very small peptides). To evaluate the pertinence of our proteomic dataset compared to the set of mitochondrial proteins already seen by coworkers, we selected 16 studies dealing with yeast mitochondria published between 1997 and 2014^{23,33,35,48-60}. Accession numbers were homogeneously converted and aligned: many proteins thus appeared to be well-characterized by the community as seen by the frequency of observation, whereas a sub-set of 23 proteins was only identified in our study. Among them, an intriguing protein named ENO1 was further investigated because it was only identified in the CESI-MS/MS dataset with 41 MS/MS spectra. Enolases are essential glycolytic enzymes that catalyse the interconversion of 2-phosphoglycerate to phosphoenolpyruvate. They are present under two isoforms, named A and B, in *Saccharomyces cerevisiae*⁶¹. It is very important to describe a biological sample as exhaustively as possible, thus including isoforms characterization by mass spectrometry: this is possible by the identification of discriminant proteotypic peptides matching on each isoform sequence. Figure 5-A displays the number of distinct peptides identified for each enolase isoform using either the CE or LC approach: isoform B (ENO2) was covered by a higher number of peptides even if the 2 proteins have slightly the same

431 molecular weight. 95% of identity is achieved on 495 residues when aligning the two protein
432 sequences with Blast-P (98% of positive matches, Figure 5-B), rendering these isoforms
433 difficult to distinguish. While investigating the detailed sequences of the matching peptides
434 using the Sequence viewer from Proteinscape software, we can conclude that ENO1 (isoform
435 A) was not validated in the LC dataset because its 6 peptides were an exact sub-set from the 8
436 peptides matching on the ENO2 sequence (ENO1 is a so-called sub-set protein). Whereas in
437 the CE dataset, 2 proteotypic peptides were identified for the ENO1 sequence and 8
438 proteotypic peptides were specifically matching on the ENO2 sequence (Figure 5-C). To note,
439 MS/MS fragmentation spectra of good quality were obtained for these proteotypic peptides
440 (Figure 5-D) and allow to distinguish two peptide sequences with close amino acids
441 composition. Likewise, the CE results also allow to observe two yeast glyceraldehyde-3-
442 phosphate dehydrogenase isoforms (G3P2 and G3P3). The CESI-MS/MS injection was
443 finally the only analysis in which we were able to distinguish the two yeast enolase isoforms
444 with a good confidence. Moreover, we also clustered the protein identifications in terms of
445 protein families, to see if specific families were identified by one of the two approaches. The
446 CESI-MS/MS dataset was seen to increase the number of members from four protein families,
447 related to mitochondrial protein import (TIM), ergosterol biosynthetic process (ERG),
448 mitochondrial electron transport (QCR) and ATP synthesis (ATP) (Supplemental Figure S5-
449 A). Proteins belonging to the TIM, QCR and ATP families are located in the mitochondrial
450 inner membrane whereas proteins from the QCR family are related to cell membrane, where
451 sterols are targeted. Given the importance of membrane proteins in various cellular processes,
452 as well as their roles in diseases, it is important that this class of proteins be better studied.
453 But identifying and characterizing proteins embedded in membranes still remains a challenge
454 in proteomics due to difficulties during the solubilization step. With the complementary
455 identifications obtained by combining orthogonal separative techniques, a better chance is

given to membrane proteins and therefore hydrophobic related peptides to be detected and identified by MS. Exploring the 2 datasets with the same type of clusterization also allows to identify protein families that are covered by CESI-MS/MS and nanoLC-MS/MS in an equivalent manner, with members specifically identified by only 1 of the 2 techniques: this is the case for the 40S and 60S ribosomal proteins family (Supplemental Figure S5-B).

Conclusion

To compare correctly datasets and to assess an eventual complementarity between CESI-MS/MS and nanoLC-MS/MS techniques, we used a large panel of peptide properties: MW, m/z , pI, Mascot Score, % peptides per protein and % sequence coverage. Based on the results presented in this study, when a low quantity of yeast mitochondrial extract is analyzed, CESI-MS/MS enables the identification of more peptides than nanoLC-MS/MS, and thus a higher protein sequence coverage. This improvement is consistent with the actual proteomics guidelines, as drastic procedures must be employed to assess the maximum of confidency on protein identification. Moreover, one of the more visible trend is concerning the peptide metric related to the MW of the observed peptides: indeed, more peptides having a molecular weight above 2000 Da have been detected by CESI-MS/MS, whereas only some of them falling into the same mass range have been identified by nanoLC-MS/MS. In parallel to the increase of the number of large peptides, the peptide charaterizing metric concerning the pI distribution was also investigated: our results indicate that more extreme pI values ($pI < 4$ and $pI > 10$) are covered by CESI-MS/MS rather than nanoLC-MS/MS. These observations are leaving the door open for the detection of peptides carrying particular biophysical properties, like phosphopeptides, without the need of enrichment techniques. In this study, the evaluation of the overall spectral quality from the CESI-MS/MS and nanoLC-MS/MS data was performed by considering the Mascot score distributions and the Paragon confidence

thresholds. Results are suggesting that increasing the sequence coverage has no detrimental impact on spectrum quality.

A challenging area of research in proteomics concerns isoforms characterization. There are several explanations for protein isoforms, among which multiple gene copies (allele variation), alternative splicing, post-translational modifications (PTM) or degradation products. In this study, we demonstrated that the 2 yeast enolase isoenzymes were both characterized in the CESI-MS/MS dataset. The observation of discriminant proteotypic peptides is facilitated when a high number of precursors with high-quality MS/MS spectra are generated. Traditional approaches are combining different proteolytic enzymes or are benefiting from the integration of bottom-up proteomics with top-down and middle-down approaches. A combination of orthogonal separative techniques, coupled online to the mass spectrometer, also appears to be a good alternative to decipher a complex proteome.

The evaluation of both techniques using a large panel of peptide metrics allow us to say that they possess complementary properties for peptide and protein identification in mitochondria isolated from cultured *S. cerevisiae*. Moreover, the power of integrating two orthogonal separative technologies and two protein classification systems offers new promising opportunities to researchers working in the mitochondrial field of yeasts and other organisms, and more broadly in subcellular proteomics.

Acknowledgments

Authors would like to thank Sciex separations Inc. for lending a CESI8000, Dr. M. Anselme and Dr. Stephen Lock from Sciex Inc. for their support. This work was supported by the “Laboratoires d'excellence” (LABEX) NetRNA grant ANR-10-LABX-36 (PR) in the frame of “Programme d'Investissements d'Avenir.”

References

1. Aebersold, R., Mann, M.; Mass spectrometry-based proteomics; *Nature*, (2003); 422: 198-207.
2. Smith, R.D., Olivares, J.A., Nguyen, N.T., Udseth, H.R.; Capillary zone electrophoresis mass-spectrometry using an electrospray ionisation interface; *Analytical Chemistry*, (1988); 60: 436-441.
3. Mann, M., Wilm, M.; Electrospray mass-spectrometry for protein characterization; *Trends in Biochemical Science*, (1995); 20: 219-224.
4. Ramautar, R., Heemskerk, A.A.M., Hensbergen, P.J., Deelder, A.M., Busnel, J.M., Mayboroda, O.A.; CE-MS for proteomics: Advances in interface development and application; *Journal of Proteomics*, (2012); 75: 3814-3828.
5. Hjerten, S.; High-performance electrophoresis - The elcctrophoretic conterpart of high-performance liquid-chromatography; *Journal of Chromatography*, (1983); 270: 1-6.
6. Jorgenson, J.W., Lukacs, K.D.; Zone electrophoresis in open-tubular glass-capillaries; *Analytical Chemistry*, (1981); 53: 1298-1302.
7. Busnel, J.M., Schoenmaker, B., Ramautar, R., Carrasco-Pancorbo, A., Ratnayake, C., Feitelson, J.S., Chapman, J.D., Deelder, A.M., Mayboroda, O.A.; High Capacity Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry: Coupling a Porous Sheathless Interface with Transient-Isotachopheresis; *Analytical Chemistry*, (2010); 82: 9476-9483.
8. Gahoual, R., Busnel, J.M., Wolff, P., Francois, Y.N., Leize-Wagner, E.; Novel sheathless CE-MS interface as an original and powerful infusion platform for nanoESI study: from intact proteins to high molecular mass noncovalent complexes; *Analytical and Bioanalytical Chemistry*, (2014); 406: 1029-1038.
9. Haselberg, R., de Jong, G.J., Somsen, G.W.; Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins; *Analytical Chemistry*, (2013); 85: 2289-2296.
10. Heemskerk, A.A.M., Busnel, J.M., Schoenmaker, B., Derks, R.J.E., Klychnikov, O., Hensbergen, P.J., Deelder, A.M., Mayboroda, O.A.; Ultra-Low Flow Electrospray Ionization-Mass Spectrometry for Improved Ionization Efficiency in Phosphoproteomics; *Analytical Chemistry*, (2012); 84: 4552-4559.
11. Geiger, M., Hogerton, A.L., Bowser, M.T.; Capillary Electrophoresis; *Analytical Chemistry*, (2012); 84: 577-596.
12. Fonslow, B.R., Yates, J.R.; Capillary electrophoresis applied to proteomic analysis; *Journal of Separation Science*, (2009); 32: 1175-1188.
13. Haselberg, R., de Jong, G.J., Somsen, G.W.; Capillary electrophoresis-mass spectrometry for the analysis of intact proteins 2007-2010; *Electrophoresis*, (2011); 32: 66-82.
14. Gahoual, R., Burr, A., Busnel, J.M., Kuhn, L., Hammann, P., Beck, A., Francois, Y.N., Leize-Wagner, E.; Rapid and multi-level characterization of trastuzumab using sheathless capillary electrophoresis-tandem mass spectrometry; *Mabs*, (2013); 5: 479-490.
15. Wojcik, R., Dada, O.O., Sadilek, M., Dovichi, N.J.; Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis; *Rapid Communications in Mass Sepctrometry*, (2010); 24: 2554-2560.

- 560 16. Zhu, G.J., Sun, L.L., Yang, P., Dovichi, N.J.; On-line amino acid-based capillary isoelectric
561 focusing-ESI-MS/MS for protein digests analysis; *Analytica Chimica Acta*, (2012); 750: 207-211.
- 562 17. Sun, L.L., Zhu, G.J., Li, Y.H., Wojcik, R., Yang, P., Dovichi, N.J.; CZE-ESI-MS/MS system for
563 analysis of subnanogram amounts of tryptic digests of a cellular homogenate; *Proteomics*, (2012); 12:
564 3013-3019.
- 565 18. Faserl, K., Sarg, B., Kremser, L., Lindner, H.H.; Optimization and Evaluation of a Sheathless
566 Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry Platform for Peptide Analysis:
567 Comparison to Liquid Chromatography-Electrospray Ionization Mass Spectrometry; *Analytical*
568 *Chemistry*, (2011); 83: 7297-7305.
- 569 19. Sun, L.L., Zhu, G.J., Zhao, Y.M., Yan, X.J., Mou, S., Dovichi, N.J.; Ultrasensitive and Fast
570 Bottom-up Analysis of Femtogram Amounts of Complex Proteome Digests; *Angewandte Chemie-*
571 *International Edition*, (2013); 52: 13661-13664.
- 572 20. Sun, L., Hebert, A.S., Yan, X., Zhao, Y., Westphall, M.S., Rush, M.J.P., Zhu, G., Champion,
573 M.M., Coon, J.J., Dovichi, N.J.; Over 10 000 Peptide Identifications from the HeLa Proteome by Using
574 Single-Shot Capillary Zone Electrophoresis Combined with Tandem Mass Spectrometry; *Angewandte*
575 *Chemie-International Edition*, (2014); 53: 13931-13933.
- 576 21. Han, X., Wang, Y., Aslanian, A., Bern, M., Lavalley-Adam, M., Yates, J.R.; Sheathless Capillary
577 Electrophoresis-Tandem Mass Spectrometry for Top-Down Characterization of *Pyrococcus furiosus*
578 Proteins on a Proteome Scale; *Analytical Chemistry*, (2014); 86: 11006-11012.
- 579 22. Faserl, K., Kremser, L., Müller, M., Teis, D., Lindner, H.H.; Quantitative Proteomics Using
580 Ultralow Flow Capillary Electrophoresis–Mass Spectrometry; *Analytical Chemistry*, (2015); 87: 4633-
581 4640.
- 582 23. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schonfisch, B.,
583 Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N., Meisinger, C.; The proteome of
584 *Saccharomyces cerevisiae* mitochondria; *Proceedings of the National Academy of Sciences of the*
585 *United States of America*, (2003); 100: 13207-13212.
- 586 24. Muller, B., Grossniklaus, U.; Model organisms - A historical perspective; *Journal of*
587 *Proteomics*, (2010); 73: 2054-2063.
- 588 25. Castrillo, J.O., Oliver, S.G.; Yeast as a touchstone in post-genomic research: Strategies for
589 integrative analysis in functional genomics; *Journal of Biochemistry and Molecular Biology*, (2004);
590 37: 93-106.
- 591 26. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F.,
592 Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin,
593 H., Oliver, S.G.; Life with 6000 genes; *Science*, (1996); 274: 546-552.
- 594 27. Pereira, C.V., Moreira, A.C., Pereira, S.P., Machado, N.G., Carvalho, F.S., Sardao, V.A., Oliveira,
595 P.J.; Investigating drug-induced mitochondrial toxicity: a biosensor to increase drug safety?; *Current*
596 *Drug Safety*, (2009); 4: 34-54.
- 597 28. Benard, G., Rossignol, R.; Ultrastructure of the mitochondrion and its bearing on function and
598 bioenergetics; *Antioxidants & Redox Signaling*, (2008); 10: 1313-1342.
- 599 29. Yi, M.Q., Weaver, D., Hajnoczky, G.; Control of mitochondrial motility and distribution by the
600 calcium signal: a homeostatic circuit; *Journal of Cell Biology* (2004); 167: 661-672.
- 601 30. Zamsami, N., Kroemer, G.; Apoptosis-condensed matter in cell death; *Nature*, (1999); 401:
602 127-128.
- 603 31. Moyle, J., Mitchell, P.; Active-inactive state transitions of mitochondrial atpase molecules
604 influenced by Mg^{2+} , anion and aurovertin; *FEBS Letters*, (1975); 56: 55-61.
- 605 32. Gaucher, S.P., Taylor, S.W., Fahy, E., Zhang, B., Warnock, D.E., Ghosh, S.S., Gibson, B.W.;
606 Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid
607 chromatography coupled with tandem mass spectrometry; *Journal of Proteome Research*, (2004); 3:
608 495-505.
- 609 33. Prokisch, H., Scharfe, C., Camp, D.G., Xiao, W.Z., David, L., Andreoli, C., Monroe, M.E., Moore,
610 R.J., Gritsenko, M.A., Kozany, C., Hixson, K.K., Mottaz, H.M., Zischka, H., Ueffing, M., Herman, Z.S.,

611 Davis, R.W., Meitinger, T., Oefner, P.J., Smith, R.D., Steinmetz, L.M.; Integrative analysis of the
612 mitochondrial proteome in yeast; *PLoS Biology*, (2004); 2: 795-804.

613 34. Meisinger, C., Sickmann, A., Pfanner, N.; The mitochondrial proteome: From inventory to
614 function; *Cell*, (2008); 134: 22-24.

615 35. Reinders, J., Zahedi, R.P., Pfanner, N., Meisinger, C., Sickmann, A.; The complete yeast
616 mitochondrial proteome: Multidimensional separation techniques for mitochondrial proteomics;
617 *Journal of Proteome Research*, (2006); 5: 1543-1554.

618 36. Tang, K.Q., Page, J.S., Smith, R.D.; Charge competition and the linear dynamic range of
619 detection in electrospray ionization mass spectrometry; *Journal of the American Society for Mass
620 Spectrometry*, (2004); 15: 1416-1423.

621 37. Zhu, G.J., Sun, L.L., Yan, X.J., Dovichi, N.J.; Single-Shot Proteomics Using Capillary Zone
622 Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry with Production of More than 1
623 250 Escherichia coli Peptide Identifications in a 50 min Separation; *Analytical Chemistry*, (2013); 85:
624 2569-2573.

625 38. Moini, M.; Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to
626 mass spectrometry using a porous tip; *Analytical Chemistry* (2007); 79: 4241-4246.

627 39. Wang, Y., Fonslow, B.R., Wong, C.C.L., Nakorchevsky, A., Yates, J.R.; Improving the
628 Comprehensiveness and Sensitivity of Sheath less Capillary Electrophoresis-Tandem Mass
629 Spectrometry for Proteomic Analysis; *Analytical Chemistry* (2012); 84: 8505-8513.

630 40. Gahoual, R., Busnel, J.M., Beck, A., François, Y.N., Leize-Wagner, E.; Full Antibody Primary
631 Structure and Microvariant Characterization in a Single Injection Using Transient Isotachophoresis
632 and Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry; *Analytical Chemistry*, (2014);
633 86: 9074-9081.

634 41. Gahoual, R., Biacchi, M., Chicher, J., Kuhn, L., Hammann, P., Beck, A., Leize-Wagner, E.,
635 Francois, Y.N.; Monoclonal antibodies biosimilarity assessment using transient isotachophoresis
636 capillary zone electrophoresis-tandem mass spectrometry; *Mabs*, (2014); 6: 1464-1473.

637 42. Sarg, B., Faserl, K., Kremser, L., Halfinger, B., Sebastiano, R., Lindner, H.H.; Comparing and
638 Combining Capillary Electrophoresis Electrospray Ionization Mass Spectrometry and Nano-Liquid
639 Chromatography Electrospray Ionization Mass Spectrometry for the Characterization of Post-
640 translationally Modified Histones; *Molecular & Cellular Proteomics*, (2013); 12: 2640-2656.

641 43. Knight, C.G., Kassen, R., Hebestreit, H., Rainey, P.B.; Global analysis of predicted proteomes:
642 Functional adaptation of physical properties; *Proceedings of the National Academy of Sciences of the
643 United States of America*, (2004); 101: 8390-8395.

644 44. Dreger, M.; Subcellular proteomics *Mass Spectrometry Review*, (2003); 22: 27-56.

645 45. L Lilley, K.S., Dupree, P.; Mant organelle proteomics; *Current Opinion in Plant Biology*, (2007);
646 10: 594-599.

647 46. Mi, H.Y., Muruganujan, A., Thomas, P.D.; PANTHER in 2013: modeling the evolution of gene
648 function, and other gene attributes, in the context of phylogenetic trees; *Nucleic Acids Research*,
649 (2013); 41: D377-D386.

650 47. Chou, K.C., Wu, Z.C., Xiao, X.A.; iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular
651 Localization of Singleplex and Multiplex Eukaryotic Proteins; *Plos One*, (2011); 6: 146-153.

652 48. Renvoise, M., Bonhomme, L., Davanture, M., Valot, B., Zivy, M., Lemaire, C.; Quantitative
653 variations of the mitochondrial proteome and phosphoproteome during fermentative and
654 respiratory growth in *Saccharomyces cerevisiae*; *Journal of Proteomics*, (2014); 106: 140-150.

655 49. Ben-Menachem, R., Tal, M., Shadur, T., Pines, O.; A third of the yeast mitochondrial
656 proteome is dual localized: A question of evolution; *Proteomics*, (2011); 11: 4468-4476.

657 50. Zahedi, R.P., Sickmann, A., Boehm, A.M., Winkler, C., Zufall, N., Schonfisch, B., Guiard, B.,
658 Pfanner, N., Meisinger, C.; Proteomic analysis of the yeast mitochondrial outer membrane reveals
659 accumulation of a subclass of preproteins; *Molecular Biology of the Cell*, (2006); 17: 1436-1450.

660 51. Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O'Shea, E.K.;
661 Global analysis of protein localization in budding yeast; *Nature*, (2003); 425: 686-691.

52. Ohlmeier, S., Kastaniotis, A.J., Hiltunen, J.K., Bergmann, U.; The yeast mitochondrial proteome, a study of fermentative and respiratory growth; *Journal of Biological Chemistry*, (2004); 279: 3956-3979.
53. Reijans, M., Lascaris, R., Groeneger, A.O., Wittenberg, A., Wesselink, E., van Oeveren, J., de Wit, E., Boorsma, A., Voetdijk, B., van der Spek, H., Grivell, L.A., Simons, G.; Quantitative comparison of cDNA-AFLP, microarrays, and GeneChip expression data in *Saccharomyces cerevisiae*; *Genomics*, (2003); 82: 606-618.
54. Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., Jacq, C.; Genome-wide analysis of mRNAs targeted to yeast mitochondria; *Embo Reports*, (2002); 3: 159-164.
55. Snyder, M., Kumar, A.; Yeast genomics: past, present, and future promise; *Functional & Integrative Genomics*, (2002); 2: 135-137.
56. Dimmer, K.S., Fritz, S., Fuchs, F., Messerschmitt, M., Weinbach, N., Neupert, W., Westermann, B.; Genetic basis of mitochondrial function and morphology in *Saccharomyces cerevisiae*; *Molecular Biology of the Cell*, (2002); 13: 847-853.
57. Steinmetz, L.M., Scharfe, C., Deutschbauer, A.M., Mokranjac, D., Herman, Z.S., Jones, T., Chu, A.M., Giaever, G., Prokisch, H., Oefner, P.J., Davis, R.W.; Systematic screen for human disease genes in yeast; *Nature Genetics*, (2002); 31: 400-404.
58. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P.; Comparative assessment of large-scale data sets of protein-protein interactions; *Nature*, (2002); 417: 399-403.
59. Pflieger, D., Le Caer, J.P., Lemaire, C., Bernard, B.A., Dujardin, G., Rossier, J.; Systematic identification of mitochondrial proteins by LC-MS/MS; *Analytical Chemistry* (2002); 74: 2400-2406.
60. DeRisi, J.L., Iyer V.R., Brown, P.O.; Exploring the metabolic and genetic control of gene expression on a genomic scale; *Science*, (1997); 278: 680-686.
61. Brandina, I., Graham, J., Lemaitre-Guillier, C., Entelis, N., Krashennikov, I., Sweetlove, L., Tarassov, I., Martin, R.P.; Enolase takes part in a macromolecular complex associated to mitochondria in yeast; *Biochimica et Biophysica Acta*, (2006); 1757: 1217-1228.

Captions

Figure 1. MS data obtained on 100ng of yeast mitochondrial extract. (A) NanoLC-MS/MS injection: heat map, where the mass-to-charge ratio of the peptides (Da) is plotted against the LC elution time (min). Each blue dot is related to a single MS/MS fragmentation spectrum. (B) CESI-MS/MS injection: heat map, where the mass-to-charge ratio of the peptides (Da) is plotted against the CE migration time (min). (C) Number of proteins and peptides identified in either the nanoLC-MS/MS (red) or the CESI-MS/MS (blue) injection.

Figure 2. Comparison of peptides molecular weight (A), mass-to-charge (m/z) ratio (B), pI (C) and Mascot score (D) by means of CESI-MS/MS (red) and nanoLC-MS/MS (blue). Data are obtained from the analysis of 100ng of a mitochondrial yeast tryptic digest on both types of coupling.

Figure 3. Effect on the global protein coverage when comparing the CESI-MS/MS (blue) and the nanoLC-MS/MS (red) datasets. (A) Proteins are ranked according to the total number of MS/MS spectra matching to each sequence (log scale) and 5 proteins, distributed all along the dynamic range, were further investigated. (B) Comparison of the sequence coverage from the 5 previously chosen proteins: shared and specific peptides are indicated in each Venn diagram, as well as the total sequence coverage increase (in %). (C) Boxplot showing the sequence coverage distributions for both datasets when considering all the identified proteins.

Figure 4. Venn diagrams indicating the distribution of mitochondrial proteins identified with Panther and iLoc-Euk databases using nanoLC-MS/MS and CESI-MS/MS approaches. Distribution of proteins between mitochondrion and other sub-cellular compartments using the nanoLC-MS/MS dataset (A) or the CESI-MS/MS dataset (B). (C) displays the overlap between the global protein dataset and the mitochondrial proteins using the two approaches.

Figure 5. How is it possible to distinguish 2 protein isoforms using discriminant peptides and an increased sequence coverage: ENO1 and ENO2 proteins as an example. (A) Comparison of the sequence coverage for each isoform: shared and specific peptides are indicated in the Venn diagrams, as well as the total sequence coverage increase (in %). (B) BLAST-P alignment between both ENO isoforms. (C) Detailed sequence coverage on both isoforms obtained by either the CESI-MS/MS or the nanoLC-MS/MS analysis: identified amino acids are shown with a red label, whereas the grey box below the corresponding peptide is indicating the MS/MS spectrum quality (a red box is attributed under each amino acid if it was seen with either a b- or a y-ion, represented in the upper and lower section of the grey box respectively). Proteotypic peptides allowing the distinction between both isoforms in each dataset are indicated with purple boxes. (D) Focus on 2 proteotypic peptides matching on ENO1 and ENO2 sequences: a difference of 2 amino acids in the composition of a peptide is easily distinguishable when inspecting the MS/MS fragmentation spectra.